Moderate Interrater Reliability in the Diagnosis of Scaphoid Waist Fractures

Rasmus Wejnold Jørgensen, MD[1] Claus Hjorth Jensen, MD[1]

[1] Department of Orthopedics, Hand Clinic, Herlev-Gentofte University Hospital of Copenhagen, Copenhagen, Denmark

J Wrist Surg 2019;8:104–107.

Address for correspondence Rasmus Wejnold Jørgensen, MD, Department of Orthopedics, Hand Clinic, Herlev-Gentofte University Hospital of Copenhagen, Copenhagen, Denmark (e-mail: Rasmus.Wejnold.Joergensen@regionh.dk).

Abstract

Background Conventional radiographs have been shown to yield unreliable results in classifying scaphoid fractures. Computed tomography (CT) has been claimed to be the tool of choice in determining the treatment as well as fracture displacement.

Purpose The purpose of the study was to examine the interrater reliability and intrarater reproducibility in the decision-making of the treatment of scaphoid waist fractures.

Patients and Methods Fifty-one CT scans of scaphoid waist fractures were utilized. Seven orthopaedic surgeons with a particular interest in hand surgery independently scrutinized the scans classifying each in undisplaced, < 2 mm displaced, or > 2 mm displaced, and suggested a treatment of immobilization in cast or screw fixation. The Fleiss' and Cohen's kappa values using SPSS (Statistical Package for Social Science) version 24 were calculated and interpreted according to Landis and Koch.

Results The kappa value representing interrater reliability when choosing between operative or nonoperative treatment was 0.58. Interrater reliability of the distinction between < 2 mm displaced or > 2 mm displaced fractures was 0.61. On average 79.5% of the fractures were suggested treated nonoperatively and 20.5% operatively. Overall, intrarater reproducibility was 0.75 when classifying between < 2 mm displaced or > 2 mm displaced fractures. When choosing between operative or nonoperative treatment, intrarater reproducibility was 0.69.

Conclusion Moderate interrater reliability was found when choosing between nonoperative and operative treatment. The use of CT showed substantial reliability in the distinction between < 2 mm displaced and > 2 mm displaced fractures. Intrarater reproducibility was substantial when classifying between < 2 mm displaced and > 2 mm displaced fractures as well as when choosing between operative or nonoperative treatment.

Level of Evidence This is a Level III study.

Keywords
- interrater reliability
- intrarater reproducibility
- scaphoid fracture
- computed tomography

Scaphoid fractures and their treatment have been discussed over several years. Nonunion of the scaphoid leads to difficult treatment, pain, and potential development of arthritis. Advantages with surgical management have been proposed as the union rate seems to be higher and the recovery is faster. However, possible complications when choosing surgery are scar-related complications, prominent hardware, chronic regional pain syndrome, and infections. Complications following cast treatment are rare, but nonunion rates seem higher.[1] Correct diagnosis and choice of treatment are crucial in the management of scaphoid waist fractures.

Apart from the clinical examination, several imaging techniques have been used such as conventional radiographs, magnetic resonance imaging (MRI), and computed tomography (CT). Conventional radiographs have been shown to be unreliable in classifying scaphoid fractures as well as in determining

received
May 4, 2018
accepted
August 2, 2018
published online
October 18, 2018

DOI https://doi.org/10.1055/s-0038-1673335.
ISSN 2163-3916.

union.[2,3] Interrater reliability is classified as poor when relying on conventional radiographs.[4,5] Gadolinium MRI has been proposed as the golden standard in the evaluation of blood supply to the proximal pole of the fracture and can detect the presence of avascular necrosis as well as healing.[5] The displacement of the fracture has been proposed as the determining factor for healing of scaphoid waist fractures and CT has been claimed to be the tool of choice in determining the degree of displacement.[3] Some studies have examined the use of CT for diagnosing union or nonunion and determining fracture healing and have shown moderate to substantial interrater reliability.[2,6,7] To our knowledge, no studies have reported on intrarater reproducibility and interrater reliability when using CT for classification of fractures and more importantly for choosing the best treatment of scaphoid waist fractures (►**Figs 1** and **2**).

The purpose of the present study was, therefore, to examine the interrater reliability and intrarater reproducibility of the classification of scaphoid waist fractures as well as the choice of treatment. The rater population consisted of specialists of orthopaedic surgery with a particular interest in hand surgery.

Patients and Methods

From 2009 to 2014, CT scans were routinely obtained in patients with fractures of the carpal scaphoid shown on conventional radiographs. Among 186 positive CT scans, 116 were excluded as the fracture was more than 4 weeks old or due to incomplete dataset of the CT. Among the remaining 70 scans, we identified 51 scaphoid waist fractures. Sagittal and coronal planes of the long axis of the scaphoid were available in all scans. The rater population consisted of seven orthopaedic surgeons with a particular interest in hand surgery. Each rater independently scrutinized the scans twice, classifying each in undisplaced, < 2 mm displaced, or > 2 mm displaced and suggested a treatment of 4 weeks of immobilization in a cast, 8 to 12 weeks of immobilization in a cast, or open reduction and internal fixation (ORIF), as proposed by Davis et al. All CT scans were anonymous and the raters were blinded to the patient history. Raters were aware that this was a study and that they were being compared both in terms of intrarater reproducibility and interrater reliability. Raters were blinded to each other's results. The time between the first and second rating was 1 week. All ratings were performed using computer screens that are used in everyday clinical work and as such represent the clinical situation. Two different scanners were used: (1) Siemens Somatom Definition, 64 slices, slice thickness 0.6 and pitch 0.9 and (2) General Electric Lightspeed VCT, 64 slice, slice thickness 0.625 and a pitch of 0.53. Secondary reconstruction was made on both machines with 2 mm slice thickness and 2 mm space (►**Fig. 3**).

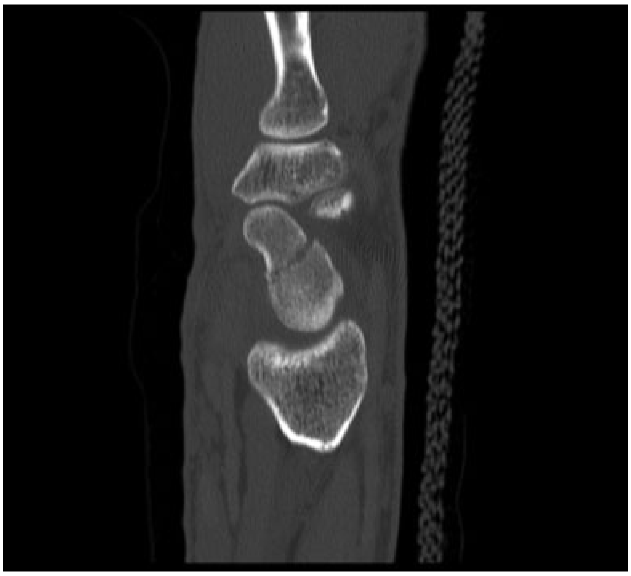
Fig. 1 Scaphoid waist fractures seen on the sagittal plane of the long axis of the scaphoid.

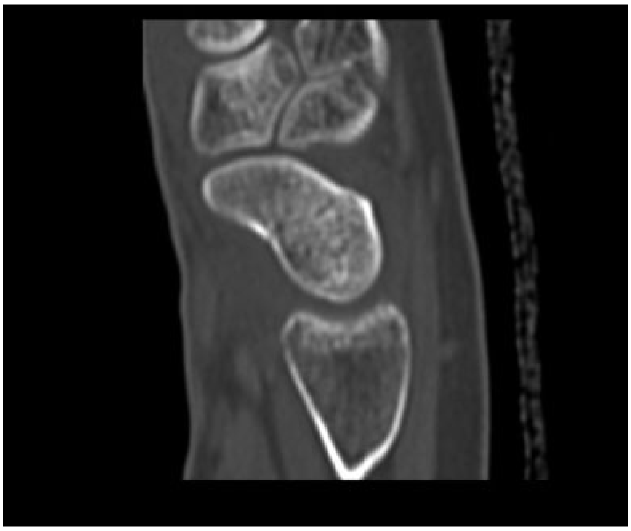
Fig. 2 Scaphoid waist fractures (undisplaced) seen on the sagittal plane of the long axis of the scaphoid.

A method of power analysis for reliability and agreement studies for categorical data does not exist. Recommendations regarding the number of observations are 40 with approximation around a mean kappa value.[8] The number of raters equally assumes a mean value with more than five raters.

Statistical analysis included the Fleiss' kappa for multiple raters and interrater reliability. An overall intrarater reproducibility was calculated using Cohen's Kappa. Overall intrarater reproducibility was calculated based on the average of ratings and not of a single rater. Kappa values were interpreted as described by Landis and Koch:[9] Kappa values of 0.01 to 0.20 indicate slight agreement, 0.21 to 0.40 fair agreement, 0.41 to 0.60 moderate agreement, 0.61 to 0.80 substantial agreement, and more than 0.80 almost perfect agreement. Zero indicates no agreement beyond that expected owing to chance alone: -1.00 means total disagreement and 1.00 represents perfect agreement. SPSS software version 24 was used.

Based on the first round of ratings, demographics of the fractures are as seen in ►**Table 1**.

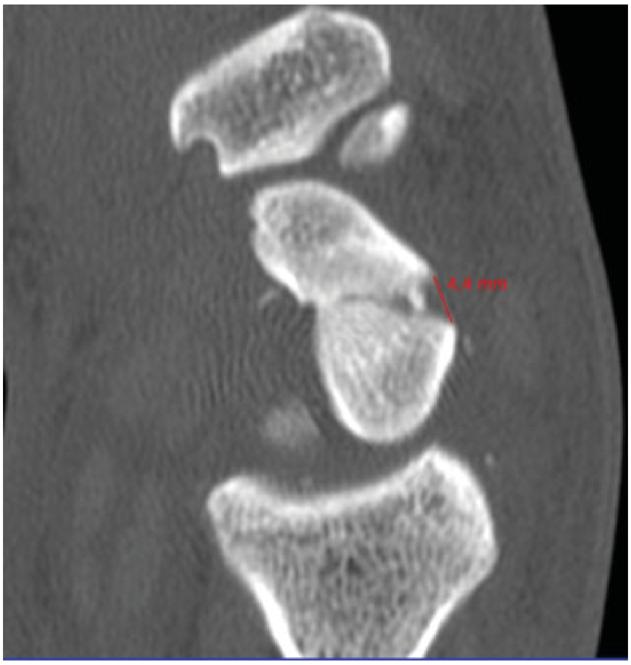

Fig. 3 Example of fracture with > 2 mm displacement.

Results

The kappa value representing interrater reliability when choosing between operative or nonoperative treatment was 0.58 ($p < 0.001$; 95% CI [confidence interval]: 0.51–0.65) interpreted as moderate agreement by Landis and Koch.[9] The interrater reliability of the classification in the categories < 2 mm displaced or > 2 mm displaced was 0.61 ($p < 0.001$; 95% CI: 0.55–0.67) interpreted as substantial agreement.[9] All kappa values increased as expected, when categories were simplified and treatment options were few. Results are shown in ►**Table 2**.

Average intrarater reproducibility with a kappa value of 0.75 ($p < 0.001$), when classifying between < 2 mm displaced or > 2 mm displaced fractures, was found (substantial agreement). Intrarater reproducibility when choosing between operative or nonoperative treatment was 0.69 ($p < 0.001$; substantial agreement). Results are shown in ►**Table 3**.

Table 1 Demographics of fractures based on first round average ratings

Demographic	%
Classification	
Undisplaced	38.5
< 2 mm displaced	41
> 2 mm displaced	20.5
Treatment	
Casting	79.5
Open reduction internal fixation	20.5

Table 2 Interrater reliability

Category	Kappa value	*p*-Value	Strength of agreement
Classification			
Undisplaced, < 2 mm displaced, > 2 mm displaced	0.502	< 0.001	Moderate
< 2 mm displaced, >2 mm displaced	0.610	< 0.001	Substantial
Treatment			
Casting 4 wk, casting 8–12 wk, open reduction internal fixation	0.397	< 0.001	Fair
Casting, open rReduction internal fixation	0.578	< 0.001	Moderate

Discussion

The diagnosis and choice of treatment for scaphoid waist fractures can be challenging as illustrated by the low interrater reliability and reproducibility examined in other studies. Apart from the clinical examination, several imaging techniques are available when examining a scaphoid fracture. Conventional radiographs have been shown to yield unreliable results in classifying scaphoid fractures as well as when determining union.[2,3] Interrater reliability was classified as poor when relying on plain radiographs as described by Dias et al.[4] CT has been claimed to be the tool of choice in determining the degree of displacement.[3] Some studies have examined the use of CT for diagnosing union or nonunion and determining fracture healing and shown moderate to substantial interrater reliability.[2,6,7]

Table 3 Intrarater reproducibility

Category	Kappa value	*p*-Value	Strength of agreement
Classification			
Undisplaced, < 2 mm displaced, > 2 mm displaced	0.661	< 0.001	Substantial
< 2 mm displaced, > 2 mm displaced	0.749	< 0.001	Substantial
Treatment			
Casting 4 wk, casting 8–12 wk, open reduction internal fixation	0.647	< 0.001	Substantial
Casting, open reduction internal fixation	0.690	< 0.001	Substantial

Limitations of this study are those inherent of an interrater reliability and reproducibility study of categorical values. Power analysis in these kinds of studies are lacking and following recommendations and guidelines for study designs are the best options available at the moment.[8,10,11]

Intrarater reproducibility measures the degree of which each rater agrees with themselves. Intrarater reproducibility was substantial in this study when classifying and choosing treatment and support the use of CT for scaphoid waist fractures.

Conventional radiographs are inferior to CT in terms of interrater reliability in the classification of union/nonunion.[6] Buijze et al[6] investigated the interrater reliability when classifying union or nonunion of scaphoid waist fractures using CT and substantial agreement represented by a kappa value of 0.66 was found. de Zwart et al[7] showed moderate interrater reliability (kappa 0.51) between four radiologists, when determining the presence of scaphoid fractures in 150 patients with clinically suspected fractures. In the present study, substantial interrater reliability in the classification between < 2 mm displaced fractures and > 2 mm displaced fractures was found. Moderate interrater reliability was found when choosing between operative or nonoperative treatment. These findings seem in accord with the current literature and support the use of CT for scaphoid waist fractures.

Ethical Approval

No human or personal human data were involved in this study.

Conflict of Interest

None declared.

Acknowledgments

The authors thank Anders Klahn, Johannes Heindl, Lars Solgård, Lars Vadstrup, Dejan Susic, Robert Gvozdenovic, and Stig Jørring for their participation as raters in this study.

References

1 Grewal R, King GJ. An evidence-based approach to the management of acute scaphoid fractures. J Hand Surg Am 2009;34(04):732–734

2 Temple CL, Ross DC, Bennett JD, Garvin GJ, King GJ, Faber KJ. Comparison of sagittal computed tomography and plain film radiography in a scaphoid fracture model. J Hand Surg Am 2005;30(03):534–542

3 Davis TR. Prediction of outcome of non-operative treatment of acute scaphoid waist fracture. Ann R Coll Surg Engl 2013;95(03):171–176

4 Dias JJ, Taylor M, Thompson J, Brenkel IJ, Gregg PJ. Radiographic signs of union of scaphoid fractures. An analysis of inter-observer agreement and reproducibility. J Bone Joint Surg Br 1988;70(02): 299–301

5 Smith M, Bain GI, Turner PC, Watts AC. Review of imaging of scaphoid fractures. ANZ J Surg 2010;80(1,2):82–90

6 Buijze GA, Wijffels MM, Guitton TG, Grewal R, van Dijk CN, Ring D; Science of Variation Group. Interobserver reliability of computed tomography to diagnose scaphoid waist fracture union. J Hand Surg Am 2012;37(02):250–254

7 de Zwart AD, Beeres FJ, Kingma LM, Otoide M, Schipper IB, Rhemrev SJ. Interobserver variability among radiologists for diagnosis of scaphoid fractures by computed tomography. J Hand Surg Am 2012;37(11):2252–2256

8 Gjørup T. Reliability of diagnostic tests. Acta Obstet Gynecol Scand Suppl 1997;166:9–14

9 Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33(01):159–174

10 Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. Stat Med 1998;17(01):101–110

11 Kottner J, Audige L, Brorson S, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. Int J Nurs Stud 2011;48(06):661–671